

UNITED STATES PATENT APPLICATION

of

Alan W. Shen,

David L.A. Anson,

and

Roman Sherman

for

**COMMUNICATING MULTI-PART MESSAGES
BETWEEN CELLULAR DEVICES USING
A STANDARDIZED INTERFACE**

109260" 836E9550

WORKMAN, NYDEGGER & SEELEY
A PROFESSIONAL CORPORATION
ATTORNEYS AT LAW
1000 EAGLE GATE TOWER
60 EAST SOUTH TEMPLE
SALT LAKE CITY, UTAH 84111

BACKGROUND OF THE INVENTION

1. The Field of the Invention

[0001] The present invention relates to the field of cellular communications. Specifically, the present invention relates to methods, systems and computer program products for communicating multi-part messages between cellular devices using a standardized interface.

2. Background and Related Art

[0002] Communication is widely perceived to be an essential human need. It is generally thought that those who master the art of communication are often most likely to develop valuable relationships and otherwise expand their circle of influence in modern society. Perhaps for this reason, inventions that advance our ability to communicate are often perceived to have the largest contribution to our civilization. Thus, two of the most valued technical innovations in the modern era have been the telephone and the radio.

[0003] The advent of the telephone allowed individuals to audibly communicate in real-time over vast distances. Conventional telephones were "hard-wired" in that the telephonic device relied on a wired connection to communicate over a telephonic network with other telephonic devices. Accordingly, conventional telephones were relatively fixed to a particular location.

[0004] The advent of the radio allowed for the real-time communication of audible information without the use of wires. At its very core, conventional radio technology involves the use of an antenna for conversion of current signals into terrestrial airwave signals that are widely broadcast. A receiver within the range of the broadcast may then

convert the terrestrial airwave signals back into the current signals preparatory to rendering the audio information via a speaker.

[0005] Wireless telephonic communication merges conventional radio and telephone technologies. Specifically, a wireless telephone includes an antenna for transmitting and receiving terrestrial airwave signals. Once a connection is established, a user may speak into the telephone. The voice information is then transmitted wirelessly using a predetermined transmission frequency to a receiver that is connected to a wired telephone network such as the traditional Public Switched Telephone Network (PSTN). Once received, the information is communicated through the connection in the wired telephone network to the other telephone. If the other telephone is wireless, then a transmitter receives the information and wirelessly transmits the information to the other telephone. The other telephone may communicate information back to the original telephone in the same manner.

[0006] Prior to cellular technology, wireless telephone communications involved radio telephones in which there was typically one central antenna tower for a relatively large area. This limited the number of channels that could be maintained for this relatively large area since each channel occupied given frequency ranges. Accordingly, only a few people could communicate via radio telephones at any given time in this relatively large area. In addition, due to the large area, the telephone transmission power needed to be significant. Accordingly, the radio telephones were quite large and bulky by today's standards, and thus used by relatively few.

[0007] Wireless telephone technology became widely available for the masses with the development of cellular technology. Cellular technology involves the spatial division of telephone use regions into relatively small areas called "cells" that may typically be in the

range of ten square miles each. Each cell includes a base station for transmitting and receiving wireless signals to and from cellular telephones within the cell.

[0008] Each base station transmits and receives wireless signals using frequencies that are typically different than the frequencies used by the immediately surrounding cells, although more remote cells may indeed reuse the same frequencies. Thus, interference from surrounding cells is minimized and frequencies may be reused throughout the network, so long as the frequencies used at any given cell differ from its immediately surrounding cells. This allows for many more available channels and thus cellular networks support orders of magnitude more simultaneous calls than the earlier radio telephone networks.

[0009] In addition, since the cell areas are relatively small, the transmission power requirements are comparatively low, and thus the telephone battery size may be kept relatively small. Furthermore, with the high frequencies allocated for cellular communication, cellular antennas may be small as well. Accordingly, cellular telephones and cellular networks have emerged as the dominant means for wireless telephone communication.

[0010] Early cellular telephones were exclusively analog meaning that the telephone processed analog signals such as analog voice signals. Later, digital telephones also became available allowing for more efficient compression and encryption technologies to thereby improve the spectral efficiency associated with cellular channels. Digital telephones process information digitally and have been used not only to communicate voice information, but also to communicate text or data messages as well.

[0011] Typically, cellular networks support such text or data message communications. However, cellular networks limit the size of message fragments that may be communicated

over the cellular network. For example, Global System for Mobile communications (or “GSM”) cellular networks often offer a service called Short Message Service (SMS) in which messages up to 140 bytes are permitted. Many other cellular technologies also support similar short messaging services.

[0012] If the message exceeds a certain size, the message is divided into multiple short message fragments of limited size such that each of the short message fragments (along with any other header data) does not exceed the size allowance of the cellular network. The header data may include, for example, routing information, a message identifier that uniquely identifies the multi-part message, an indication of the number of short message fragments in the identified multi-part message, the order of the corresponding short message fragment in the multi-part message, and the like.

[0013] The process of fragmenting such multi-part messages into individual short message fragments is quite complex. For example, the permissible size of the short message fragment is not necessarily fixed, but depends on the size of the header information as well. Hereinafter, “a short message fragment” refers to a portion of text or data that was fragmented from the original larger message. A “short message” refers to a short message fragment along with accompanying header data.

[0014] For example, suppose a cellular network limits the size of short messages to 140 bytes. Although the size limit of the short message is 140 bytes, one simply cannot say that each short message fragment should be 140 bytes since the short message fragment will be accompanied by header information. If, for a given short message fragment, the ultimate size of the associated header information is 35 bytes, the absolute size limit for that short message fragment will be 105 bytes. If, for another short message

fragment, the ultimate size of the associated header information is 45 bytes, the absolute size limit for that other short message fragment is 95 bytes.

[0015] In addition, as previously implied, fragmentation requires the formation of header information to allow for proper reassembly of the various short message fragments at the receiving cellular device. Thus, when including such header information, one must consider the capabilities of the receiving cellular device to interpret the header information. Also, since cellular networks may delivery short message fragments out of order, there should be ordering information included in the header. Thus, the implementation of these fragmentation and reassembly functions often involve significant coding effort.

[0016] Conventionally, any application that offered the communication of messages over cellular networks had to individually deal with these non-trivial fragmentation and reassembly issues. Thus, each application provider needed to individual author code to address fragmentation and reassembly. Should standards for fragmentation and reassembly change or expand, each application provider would have to address the change. Therefore, what is desired are methods, systems, and computer program that relieve cellular application providers from having to author code that deals with fragmentation and reassembly of short message fragments.

SUMMARY OF THE INVENTION

[0017] Methods, systems, and computer program products are described that allow for fragmentation of a multi-part message into a number of short message fragments and transmission of such fragments over a cellular network without the transmitting application program dealing with the details required to perform such fragmentation. In addition, reassembly of a multi-part message is accomplished without the receiving application being concerned with the details involved with reassembly of short message fragments into a multi-part message.

[0018] When transmitting a multi-part message, a calling application issues a function call to a short messaging layer via a standardized interface such as an Application Program Interface (API) or user interface. The function call represents a request to transmit the message. After determining that the message is too large to transmit as a single short message, the short messaging layer divides the message into fragments of a sufficiently small size that adding header information will result in a short message that is within the size requirements of the cellular network. Each of the short messages are then transmitted across the cellular network.

[0019] If the calling application has also requested a delivery report, then the short messaging layer tracks delivery reports subsequently received for each of the short messages. Once all of the delivery reports for the short messages are received, the short messaging layer interprets the various delivery reports to determine an appropriate delivery report for the multi-part message as a whole. The short messaging layer then passes the appropriate delivery report back to the calling application.

[0020] When receiving a multi-part message, various short message fragments corresponding to the multi-part message are received by the short messaging layer. The

short messaging layer then uses the header information associated with each of the short message fragments to assemble the short message fragments into the multi-part message. The short messaging layer then passes the reassembled multi-part message to the receiving application. Although this process may occur automatically without a request, in one embodiment, this service is provided via an express request to do so received from the receiving application.

[0021] Thus, when transmitting, receiving, or tracking delivery associated with a multi-part message, the application program need not be concerned with details involved with the limited size requirements of the underlying cellular network. Instead, the application program merely requests transmission of the multi-part message, receives the multi-part message, and optionally receives a delivery report for the multi-part message as a whole. Should standards change for transmitting or receiving multi-part messages, the application program need not change. Instead, the short messaging layer may be adapted to deal with the change in standards.

[0022] Additional features and advantages of the invention will be set forth in the description which follows, and in part will be obvious from the description, or may be learned by the practice of the invention. The features and advantages of the invention may be realized and obtained by means of the instruments and combinations particularly pointed out in the appended claims. These and other features of the present invention will become more fully apparent from the following description and appended claims, or may be learned by the practice of the invention as set forth hereinafter.

BRIEF DESCRIPTION OF THE DRAWINGS

[0023] In order to describe the manner in which the above-recited and other advantages and features of the invention can be obtained, a more particular description of the invention briefly described above will be rendered by reference to specific embodiments thereof which are illustrated in the appended drawings. Understanding that these drawings depict only typical embodiments of the invention and are not therefore to be considered to be limiting of its scope, the invention will be described and explained with additional specificity and detail through the use of the accompanying drawings in which:

[0024] Figure 1 schematically illustrates a telephone network in which the principles of the present invention may be implemented.

[0025] Figure 2 illustrates the relevant hardware and software components of a telephonic device that may communicate over the cellular network of Figure 1, and that may implement the principles of the present invention.

[0026] Figure 3 illustrates an architecture that may be implemented on the telephonic device of Figure 2 and that operates to relieve the application provider from having to implement the details of fragmentation and reassembly of a multi-part message.

[0027] Figure 4 illustrates a flowchart of a method for facilitating an application sending multiple short message fragments without having the calling application implement the detailed processing required to fragment the message.

[0028] Figures 5A and 5B illustrates a more detailed flowchart for facilitating an application sending multiple short message fragments.

[0029] Figure 6 illustrates a flowchart of a method for responding to delivery reports for short messages.

[0030] Figure 7 illustrates a flowchart of a method for receiving a multi-part message in a manner that the receiving application does not have to implement the detailed processing necessary to reassemble the message.

[0031] Figure 8 illustrates a more detailed flowchart of a method for receiving a multi-part message in a manner that the receiving application does not have to implement the detailed processing necessary to reassemble the message.

[0032] Figure 9 illustrates a data structure of a tracking list used to provide a unified delivery report for a multi-part message.

FILED 03/25/03 10:56:00

WORKMAN, NYDEGGER & SEELEY
A PROFESSIONAL CORPORATION
ATTORNEYS AT LAW
1000 EAGLE GATE TOWER
60 EAST SOUTH TEMPLE
SALT LAKE CITY, UTAH 84111

DETAILED DESCRIPTION OF THE INVENTION

[0033] The present invention extends to methods, systems and computer program products for facilitating fragmentation and reassembly of multi-part messages for transmission over a cellular network without the application program having to implement the details necessary to fragment and reassemble the multi-part message. In particular, a calling application that is to transmit a multi-part message uses a standardized interface to request that a short messaging layer transmit the message. If the message is large enough, the short messaging layer fragments and transmits the message and, if requested, returns a unified delivery report for the multi-part message. On the receiving end, a short messaging layer reassembles the multi-part message and passes the assembled message to a receiving application, without the receiving application having to implement the details required to reassemble a multi-part message.

[0034] The embodiments of the present invention may comprise a special purpose or general purpose computing device including various computer hardware, as discussed in greater detail below. Embodiments within the scope of the present invention also include computer-readable media for carrying or having computer-executable instructions or data structures stored thereon. Such computer-readable media can be any available media which can be accessed by a general purpose or special purpose computer. By way of example, and not limitation, such computer-readable media can comprise physical storage media such as RAM, ROM, EEPROM, CD-ROM or other optical disk storage, magnetic disk storage or other magnetic storage devices, or any other medium which can be used to carry or store desired program code means in the form of computer-executable instructions or data structures and which can be accessed by a general purpose or special purpose computer.

[0035] When information is transferred or provided over a network or another communications connection (either hardwired, wireless, or a combination of hardwired or wireless) to a computer, the computer properly views the connection as a computer-readable medium. Thus, any such connection is properly termed a computer-readable medium. Combinations of the above should also be included within the scope of computer-readable media. Computer-executable instructions comprise, for example, instructions and data which cause a general purpose computer, special purpose computer, or special purpose processing device to perform a certain function or group of functions.

[0036] Although not required, the invention will be described in the general context of computer-executable instructions, such as program modules, being executed by computing devices. Generally, program modules include routines, programs, objects, components, data structures, and the like that perform particular tasks or implement particular abstract data types. Computer-executable instructions, associated data structures, and program modules represent examples of the program code means for executing steps of the methods disclosed herein. The particular sequence of such executable instructions or associated data structures represent examples of corresponding acts for implementing the functions described in such steps.

[0037] Those skilled in the art will appreciate that the invention may be practiced in network computing environments with many types of computer system configurations, including personal computers, hand-held devices, multi-processor systems, microprocessor-based or programmable consumer electronics, network PCs, minicomputers, mainframe computers, and the like. The invention may also be practiced in distributed computing environments where tasks are performed by local and remote processing devices that are linked (either by hardwired links, wireless links, or by a

combination of hardwired or wireless links) through a communications network. In a distributed computing environment, program modules may be located in both local and remote memory storage devices.

[0038] Figure 1 illustrates, in simplified form, a telephone network 100 in which the present invention may operate. The telephone network 100 facilitates communication between a number of telephonic devices such as the illustrated telephonic devices 101 through 106. Some of the telephonic devices are cellular devices and thus communicate over the telephonic network via a cellular network. For example, telephonic devices 101, 103 and 105 are cellular devices and communicate with over the telephone network 100 using corresponding cellular networks 111, 113 and 115.

[0039] Cellular networks are often not only capable of communicating voice information, but are also capable of communicating other types of messages such as text and data messages. However, often such cellular networks limit the size of such messages. For example, Short Message Service (SMS) in Global System for Mobile communication (GSM) is one such message service. In addition, Time Division Multiple Access (TDMA), Code Division Multiple Access (CDMA), Wireless CDMA (WCDMA), 1xRTT, 3G, UMTS and CDMA2000, among others, are cellular technologies that enable the transmission of messages of limited size.

[0040] The present invention facilitates the fragmentation and reassembly of such messages when the size of the message so warrants. Note that although each of the telephonic devices 101 through 106 are illustrated in the conventional form of a telephone, a telephonic device may be any device capable of communicating over a telephonic network.

[0044] Although the telephonic device 101 has the appearance of a mobile telephone, the unseen features of the telephonic device 101 may allow for complex and flexible general-purpose processing capabilities. For example, the telephonic device 101 also includes a processor 211 and a memory 212 that are connected to each other and to the user interface 201 via a bus 210. The memory 212 generically represents a wide variety of

volatile and/or non-volatile memories that may be employed. The particular type of memory used in the telephonic device 101 is not important to the present invention.

[0045] Program code means comprising one or more program modules may be stored in memory 212. The one of more program modules may include an operating system 213, one or more application programs 214, other program modules 215, and program data 216. The environment illustrated in Figure 2 is illustrative only, and by no means represents even a small portion of the wide variety of cellular devices in which the principles of the present invention may be implemented.

[0046] Figure 3 is an architectural diagram illustrating the interaction between various components implemented by a cellular device such as telephonic device 101 in accordance with the present invention. An application 301, which may be one of application programs 214, provides a service which requires a message be sent over a cellular network such as cellular network 111. Instead of the application 301 having to address all of the detailed requirements associated with transmission of a message over the cellular network, the application 301 issues a function call to the short messaging layer 302 requesting transmission of the message.

[0047] The short messaging layer 302 then implements a method 400 illustrated in Figure 4. Figure 4 is a flowchart of a method for facilitating an application sending multiple short message fragments without having the calling application implement the detailed processing required to fragment the message. The method 400 includes an act of the short messaging layer 302 receiving a function call from the calling application 301 via a standardized interface 304 (act 401). The function call represents a request to transmit a message over the cellular network.

[0048] In this description and in the claims, a “function call” is defined as an request for service from one software module to another whether it be an automatic as a result of the operation of the calling software module, or whether the call is in response to user input to a standardized user interface, such as a user entering a command to transmit in a command line.

[0049] The short messaging layer 302 then performs a step for transmitting the message over the cellular network in response to the function call (step 402). The step 402 may include corresponding acts 403 and 404. Specifically, the short messaging layer 403 divides the message into a number of short message fragments of limited size (act 403) and then causes each of the short message fragments to be transmitted over the cellular network (act 404) via the network driver 303 of Figure 3.

[0050] If the short messaging layer 302 received a request for a delivery report (YES in decision block 405, then the short messaging layer 302 gathers delivery reports for the short message fragments from the cellular network 111 (act 406) via the network driver 303. The various delivery reports are then interpreted to determine an appropriate delivery report for the multi-part message as a whole (act 407). The delivery report for the multi-part message is then returned to the application 301 via the standardized interface 304.

[0051] Figure 5 is a more detailed flowchart that illustrates a specific embodiment of a method 500 for facilitating an application sending multiple short message fragments without having the calling application implement the detailed processing required to fragment the message. The method begins when the short messaging layer 302 receives a function call via a standardized interface requesting transmission of a message (act 501). Act 501 of Figure 5 is analogous to act 401 of Figure 4.

[0052] Next, the short messaging layer 302 performs any processing that needs to be done on the message as a whole (act 502). Such processing may include compression of the message, wrapping the messaging in eXtensible Markup Language (XML), encryption and the like.

[0053] The short messaging layer 302 then calculates the number of bytes available for each short message fragment based on the type of message (act 503). For example, a text message may be permitted to have short messages of a certain size while Wireless Application Protocol (or “WAP”) message may have short messages of a different size.

[0054] The short messaging layer 302 then determines if the message could fit below the maximum permissible size set by the cellular network using a single message (decision block 504). If not (NO in decision block 504), then the message is treated as a multi-part message. Specifically, the short messaging layer 302 constructs header information that would assist in reassembling the corresponding short message fragments into the original message (act 505). Such header information may include, for example, an identification of the multi-part message, the number of short message fragments in the multi-part message, and the order of the associated short message fragment in the multi-part message.

[0055] This additional header information, of course, changes the allowable size of the raw short message fragment. The short messaging layer 302 then recalculates the maximum allowable size of a short message fragment (act 506) considering the newly constructing ordering header information. For example, suppose the cellular network allowed for a maximum short message size of 140 bytes given the type of message. Now suppose the ordering header information is 20 bytes. The allowable size of the short message fragment with the ordering header information would only be 120 bytes.

[0056] Suppose that the maximum short message fragment size is calculated to be 120 bytes. If the message (after performing any processing on the message as a whole in act 502) is 620 bytes, then 6 short message will be needed to transmit the complete multi-part messages. Some cellular networks or applications limit the number of short messages that may be associated with a single multi-part message. For example, GSM networks are currently limited to 255 short message fragments per multi-part message. If the number of short messages would exceed the maximum allowable (YES in decision block 507), then an appropriate error message is returned to the calling application (act 508). Otherwise (NO in decision block 507), the message is fragmented into the multiple short message fragments (act 509) using the maximum size of the fragments calculated in act 506. After the multi-part message has been fragmented in act 509, a complete short message is constructed for each short message fragment. This may include appending or prepending the appropriate header information constructed in act 505.

[0057] After the short messages have been constructed for each short message fragment in the multi-part message (i.e., after act 510), or after it has been determined that the message may be transmitted as a single short message (YES in decision block 504), each of the one or more short messages is encoded as needed in order to be transmitted over the cellular network (act 511). This functionality may be incorporated in the network driver 303 or may be instead implemented by the short messaging layer 302. After encoding (act 511) each short message may be transmitted over the cellular network (act 512) using, for example, the network driver 303.

[0058] If there is an error sending any of the short messages (YES in decision block 513), then an error is reported (act 514) and the process ends. Otherwise (NO in decision block 513), a tracking number for each corresponding short message is included in a

tracking list associated with the multi-part message (act 515). The structure of a tracking list is illustrated in Figure 9 as tracking list 900. The tracking list 900 includes a field 901 that represents the identification of the multi-part message (e.g., "A"), and fields 902 that identify each short message in the multi-part message (e.g., parts 1 through 5 of 5). The illustrated identifications in Figure 9 are human readable for clarity. The actual identification may be any binary sequence that identifies the corresponding fields to the telephonic device 101. The short messaging layer 302 then returns the tracking list to the calling application (act 516) and awaits delivery reports for each short message.

[0059] Figure 6 illustrates a flowchart of a method 600 for how the short messaging layer 302 responds to delivery reports for short messages. When a delivery report for a short message is received (act 601), the short messaging layer 302 removes the tracking number associate with the short message from the tracking list (act 602). If delivery reports for less than all of the short messages have been received (NO in decision block 603), then the process ends until further delivery reports are received.

[0060] If delivery reports for all of the short messages corresponding to the multi-part message have been received (YES in decision block 603), then delivery status for the entire multi-part message is reported (act 604). If any of the short messages were reported as not delivered, then the short messaging layer 302 may attempt to resend the fragment. If ultimately less than all of the fragments were reported as delivered, even after any reattempts, then the multi-part message delivery report may indicate an error corresponding to "message not delivered" or something similar. On the other hand, if all of the short messages were reported as delivered, then the multi-part message delivery reported may indicate a success code corresponding to "message delivered".

[0061] In this manner, the calling application 301 simply issued a function call to send a message and received a single delivery report back if requested. The calling application 301 did not need to deal with the many details involved with fragmenting the message into multiple short messages. Also, the calling application 301 did not have to be concerned with interpreting delivery reports for individual message fragments. Instead, the calling application 301 received a single report for the message. Accordingly, the calling application 301 need not be aware of anything other than that it requested a message to be sent, and that it received a message delivery report back if it was requested.

[0062] The same architecture of Figure 3 that was used to send a multi-part message may also be used to receive a multi-part message. Figure 7 illustrates a method 700 for receiving a multi-part message in a manner that a receiving application (e.g., application 301) does not have to perform the detailed processing necessary to reassemble the message. First, the short messaging layer 302 receives a number of short message fragments corresponding to a multi-part message (act 701), reassembles the number of short message fragments into the multi-part message (act 702), and passes the reassembled message to the receiving application via a structured standardized interface (act 703). The short messaging layer 302 may have performed the method 700 in response to an express request to do so received from the receiving application via the standardized interface.

[0063] Figure 8 illustrates a flowchart of a more specific method 800 for receiving a multi-part message in a manner that a receiving application does not have to perform the detailed processing necessary to reassemble the message. The method 800 is initiated when a short message is received (act 801) via, for example, the network driver 303. The short message is then decoded as needed (act 802). This decoding process is essentially the inverse of the encoding process performed in act 511 of Figure 5B.

[0064] The header information for the short message is then consulted to determine whether the short message is part of a multi-part message (decision block 803). If so (YES in decision block 803), the short message is saved to persistent memory (act 804) to protect against losing data in case of a power interruption while only part of the multi-part message is received.

[0065] Then, the header information is again consulted to determine if all of the short messages corresponding to the multi-part message have been received (decision block 805). If less than all of the short messages have been received (NO in decision block 805), then the process returns to act 801 to await the next short message. If, on the other hand, all of the short messages for the multi-part message have been received (YES in decision block 805), then the header information is used to reassemble the short message fragments into a single multi-part message (act 806).

[0066] Then, the short messaging layer 302 performs any processing that is needed on the message as a whole such as decompression, unwrapping from XML, or decryption (act 807), delivers the message to the receiving application via the standardized interface 304 (act 808), and then removes the short messages from the persistent memory (act 809).

[0067] When receiving the multi-part message, the receiving application remains unaware of the detailed associated with reassembly of the various short message fragments into the multi-part message. Instead, these services are provided by the short messaging layer 302.

[0068] Accordingly, the principles of the present invention allow for an application to send multi-part messages over a cellular network without the application having to implement the detailed processing required for fragmentation and reassembly of multi-part messages. The present invention may be embodied in other specific forms without

departing from its spirit or essential characteristics. The described embodiments are to be considered in all respects only as illustrative and not restrictive. The scope of the invention is, therefore, indicated by the appended claims rather than by the foregoing description. All changes which come within the meaning and range of equivalency of the claims are to be embraced within their scope.

[0069] What is claimed and desired to be secured by United States Letters Patent is:

WORKMAN, NYDEGGER & SEELEY
A PROFESSIONAL CORPORATION
ATTORNEYS AT LAW
1000 EAGLE GATE TOWER
60 EAST SOUTH TEMPLE
SALT LAKE CITY, UTAH 84111

FILED
MAR 22 1994
FBI - SALT LAKE CITY